JUPITER PRIME

Consulting • Training • Advisory

جوبيتر برايم للاستشارات والتدريب

كتيّب الدورة

دورة رائد أمن الذكاء الاصطناعي لفرق التقنية والمخاطر

10 – 14 مايو 2027

مدريد

info@jupiterprime.uk

jupiterprime.uk

مسجّلة في المملكة المتحدة والمملكة العربية السعودية

دورة رائد أمن الذكاء الاصطناعي لفرق التقنية والمخاطر

المرجع: 2918_116 **التاريخ:** 10 – 14 مايو 2027 **المكان:** مدريد **الرسوم:** GBP 6200

رائد أمن الذكاء الاصطناعي لفرق التقنية والمخاطر دورة تدريبية مدتها أسبوع لمحللي الأمن ومهندسي تقنية المعلومات ومسؤولي المخاطر الذين يتولون دور الرائد لأنظمة الذكاء الاصطناعي التوليدية والتنبؤية في منشآتهم، ويخرج منها كل مشارك بدليل أمن الذكاء الاصطناعي. فالمهاجمون يستهدفون اليوم النماذج والأوامر ومجموعات التدريب، بينما يستعمل الموظفون روبوتات محادثة غير معتمدة بسرعة تفوق تحديث السياسات. تقدم جوبيتر برايم للتدريب دورة رائد أمن الذكاء الاصطناعي.

لماذا يهم رائد أمن الذكاء الاصطناعي الآن

تشكل ثلاثة ضغوط دور الرائد اليوم. فالنماذج اللغوية المرتبطة بالبريد والملفات والإضافات تفتح مسارات هجوم جديدة مثل حقن الأوامر وتسريب المعلومات عبر النص المولد. ويلصق الموظفون مواد سرية في خدمات غير مرخصة، فينشأ ذكاء اصطناعي خفي لا تراه فرق الأمن. وتسأل مجالس الإدارة والمدققون عن المسؤول عن حوادث النماذج. المنشأة التي تملك روادا مدربين تكتشف المدخلات المسممة والتكاملات المسربة مبكرا، والتي تفتقدهم تعرف بالانكشاف من عملائها أو من الصحافة. وتستند الدورة إلى معيار ISO/IEC 42001 لنظام إدارة الذكاء الاصطناعي وإطار NIST AI RMF Framework NIST AI Risk Management.

المستفيدون من رائد أمن الذكاء الاصطناعي

- محللو الأمن الذين يراقبون التنبيهات ويحتاجون إلى تمييز أنماط الهجوم الخاصة بالنماذج
- مهندسو التقنية الذين يربطون النماذج اللغوية بصناديق البريد ومخازن المستندات والإضافات
- مسؤولو المخاطر الذين يحدثون سجل حالات استخدام الذكاء الاصطناعي وتصنيفات تعرضها
- قادة التوعية بأمن المعلومات الذين يوجهون الموظفين إلى الاستعمال الآمن للخدمات التوليدية
- مالكو التطبيقات الذين يعتمدون أو يرفضون خصائص الذكاء الاصطناعي الجديدة في أنظمة الأعمال

تحقق وظيفة الأمن التي ترسل محللا ومهندسا ومسؤول مخاطر معا أكبر فائدة، لأنهم يبنون صورة تهديدات موحدة ومسار تصعيد مشتركا.

قدرات رائد أمن الذكاء الاصطناعي

يغادر المشاركون وهم قادرون على تسمية مسارات الهجوم الخاصة بتعلم الآلة والنماذج اللغوية واختبارها واحتوائها.

- تقييم حالات استخدام الذكاء الاصطناعي بنموذج تهديدات يغطي المدخلات والنموذج والمخرجات والتكاملات — مالك التطبيق الذي يعتمد الإطلاق
- اختبار واجهات المحادثة ضد حقن الأوامر المباشر وغير المباشر — المهندس الذي يحصن التكامل
- تحديد ضوابط منع التسريب في الأوامر ومخازن الاسترجاع والنص المولد — مسؤول حماية البيانات
- موازنة مخاطر التسميم وسلاسل التوريد في النماذج ومجموعات البيانات الخارجية — قرار الشراء
- تصميم إجراء لاكتشاف الذكاء الاصطناعي الخفي واعتماد الخدمات المرخصة — رئيس أمن المعلومات
- إيصال قواعد الاستعمال الآمن عبر جلسات توعية موجهة — الموظفون في كل وحدة أعمال

مصطلحات رائد أمن الذكاء الاصطناعي الأساسية

- حقن الأوامر — نص مصمم يلغي تعليمات النموذج المقصودة، سواء كتب مباشرة أو أخفي في مستندات وصفحات يقرؤها النموذج
- تسميم البيانات — إفساد متعمد لبيانات التدريب أو الاسترجاع بحيث يتعلم النموذج سلوكا يختاره المهاجم
- الذكاء الاصطناعي الخفي — استعمال الموظفين لخدمات ذكاء اصطناعي دون موافقة أو اطلاع من وظائف التقنية والأمن
- استخلاص النموذج — استخراج معاملات النموذج أو تعليماته الداخلية أو سياقه السري عبر استعلامات متكررة
- حاجز الحماية — مرشح تقني أو فحص سياسات يوضع قبل النموذج أو بعده لحجب المدخلات أو المخرجات غير الآمنة
- الفريق الأحمر — اختبار هجومي منظم لنظام ذكاء اصطناعي لكشف الثغرات القابلة للاستغلال قبل المهاجمين

رائد أمن الذكاء الاصطناعي يوما بيوم

اليوم 1: سطح الهجوم على الذكاء الاصطناعي ودور الرائد

- خريطة مكونات نظام الذكاء الاصطناعي — المدخلات والنماذج والمخرجات والإضافات المتصلة
- ميثاق دور الرائد — التفويض وخطوط التصعيد وإيقاع التقارير
- مشهد التهديدات للنماذج اللغوية — ترتيب الهجمات المرصودة فعليا
- وظيفتا الحوكمة والتخطيط في إطار NIST AI RMF — تحديد المساءلة عن خطر النموذج
- أهداف الضوابط في معيار ISO/IEC 42001 — ربط مهام الأمن بنظام الإدارة

اليوم 2: هجمات الأوامر والمخرجات

- تقنيات حقن الأوامر المباشر — اختبار واجهات المحادثة قبل الإطلاق
- الحقن غير المباشر عبر المستندات ومحتوى الويب — تحصين مسار الاسترجاع
- أنماط كسر القيود وتجاوز الحواجز — اختيار طبقات الترشيح
- التعامل غير الآمن مع المخرجات — منع وصول الشيفرة والروابط المولدة إلى الأنظمة
- الصلاحيات المفرطة للإضافات — تقييد ما تستطيع الإجراءات المتصلة فعله

اليوم 3: تهديدات البيانات والنماذج وسلاسل التوريد

- تسميم بيانات التدريب والاسترجاع — فحوص سلامة المصادر المستوردة
- كشف المعلومات الحساسة — الحجب وضبط الوصول في مخازن السياق
- استخلاص النموذج وتسرب التعليمات الداخلية — حدود المعدل والمراقبة
- فحص النماذج ومجموعات البيانات الخارجية — استبيانات أمن الموردين
- ورقة نمذجة تهديدات الذكاء الاصطناعي — تقدير الاحتمال والأثر لكل حالة استخدام

اليوم 4: الذكاء الاصطناعي الخفي والمراقبة والاستجابة

- اكتشاف الذكاء الاصطناعي الخفي عبر إشارات الشبكة والمصروفات — قوائم الخدمات المرخصة
- السجلات والقياس عن بعد لتفاعلات النماذج — قواعد اكتشاف سوء الاستعمال
- دليل الاستجابة لحوادث الذكاء الاصطناعي — خطوات الاحتواء والتراجع والإبلاغ
- تصميم تمرين الفريق الأحمر — النطاق وقواعد الاشتباك والنتائج
- تخطيط جلسات التوعية — رسائل مصممة حسب الدور لوحدات الأعمال

اليوم 5: التطبيق العملي لدليل أمن الذكاء الاصطناعي

- استعراض نموذج التهديدات — حالتا روبوت محادثة مصرفي ومساعد سجلات مستشفى
- تمرين اختبار الحقن والتسريب — تطبيق أساليب اليوم 2 واليوم 3
- محاكاة حادث على الطاولة — تجربة دليل الاستجابة
- مطابقة الضوابط مع ISO/IEC 42001 وNIST AI RMF — ترتيب الفجوات
- دليل أمن الذكاء الاصطناعي — نماذج التهديدات والضوابط وأدلة الاستجابة وتقويم التوعية

تطبيق رائد أمن الذكاء الاصطناعي بعد الدورة

- إدارة جلسة نمذجة تهديدات لحالة استخدام الذكاء الاصطناعي الأكثر انكشافا وتسجيل التقديرات والضوابط في السجل المعتمد
- إطلاق حملة لاكتشاف الذكاء الاصطناعي الخفي ونشر قائمة الخدمات المرخصة مع إحاطة توعية قصيرة للموظفين

أسئلة حول رائد أمن الذكاء الاصطناعي

من هو رائد أمن الذكاء الاصطناعي ولماذا يهم هذا الدور الآن؟

رائد أمن الذكاء الاصطناعي هو الشخص الذي يرصد التهديدات الخاصة بأنظمة الذكاء الاصطناعي ويختبرها ويحتويها داخل المنشأة. ويهم الدور الآن لأن النماذج تتصل ببيانات حساسة، والموظفون يستعملون خدمات غير معتمدة أسرع من تغير السياسات.

ما الأدوار الأكثر استفادة من تدريب رائد أمن الذكاء الاصطناعي؟

محللو الأمن ومهندسو التقنية ومسؤولو المخاطر وقادة التوعية هم الأكثر استفادة. وتظهر حالات من المصارف والرعاية الصحية كيف يؤثر حقن الأوامر وتسريب المعلومات والتسميم في خدمة العملاء والسجلات وأنظمة دعم القرار.

كيف يختلف تدريب رائد أمن الذكاء الاصطناعي عن دورة عامة في الأمن السيبراني؟

يركز تدريب رائد أمن الذكاء الاصطناعي على هجمات لا توجد إلا بسبب النماذج والأوامر وبيانات التدريب. أما الدورة العامة فتغطي الشبكات والأجهزة والهويات، بينما تختبر هنا الحقن والتسميم والتسريب والذكاء الاصطناعي الخفي مباشرة.

ماذا تعود به من تدريب رائد أمن الذكاء الاصطناعي؟

تعود بدليل مصمم لحالات استخدام الذكاء الاصطناعي في منشأتك. ويضم نماذج التهديدات وخريطة الضوابط ودليل الاستجابة للحوادث وتقويم التوعية، جاهزا للعرض على قيادة الأمن.

رائد أمن الذكاء الاصطناعي: الطريق إلى الأمام

يمنح دليل أمن الذكاء الاصطناعي كل رائد نماذج تهديدات وضوابط مطابقة ودليل استجابة وتقويم توعية. وأول قرار يدعمه هو تحديد حالة الاستخدام التي تنال الضوابط والاختبار قبل الإصدار التالي، ويستعمله قادة الأمن لتوزيع المسؤوليات. وبعد أول حادث أو تمرين للفريق الأحمر تستحق تقديرات التهديد مراجعة، لأن المهاجمين يتكيفون بسرعة مع الدفاعات الجديدة.